Team Members

Alex Leonardi, Ben Harpe, Michael Fein

Problem Definition:

Recent advances in computer vision have greatly expanded the capabilities of deep learning models in many respects. Many of these capabilities arise from the use of massive datasets to build robust latent spaces to understand images more fundamentally. Modern models are able to identify features of images and generate new images based on descriptions, so one natural extension we can consider is a combination of these two tasks. In particular, there is an interesting intersection between identifying existing features in an image and modifying them based on a given text description.

We would like to be able to allow users to upload an image and modify that image automatically using text commands. Though there are models that allow us to tackle this problem already, there may be several avenues to expand the functionality and reach of these existing models, and transfer learning techniques may be useful in helping us achieve that goal.

Proposed Solution:

For this project, we're planning to base most of our work off of StyleCLIP, which has already achieved impressive results applying text-driven manipulations to images. We aim to create a model that can manipulate a wider variety of features in images and be able to deploy it and make it easily accessible in an application. Our final solution will be an app that uses our extension and optimization of StyleCLIP to allow users to type in manipulations to a wide variety of features in an image and quickly see the results.

A Rough Timeline:

We'll first spend time exploring the existing StyleCLIP model, its results, and the data used to train it. Before adding our own extensions to the model, we'll containerize the model and deploy it to the cloud to set up our infrastructure for the rest of the

project. This experience with the model and its capabilities will allow us to target shortcomings that we can fine-tune and ideal novel manipulations that we can add.

Next, we'll spend time finding new data and defining any new layers to the model we want to add to the model to improve performance and teach new manipulations through transfer learning. We'll then set up a cluster on GCP to continue training the model and monitor/optimize it with tools seen in lectures.

Finally, we'll deploy our finished model using our existing infrastructure and we'll build a simple front-end to allow users to easily upload/take pictures, query our model's API, and display the final results.

References:

StyleCLIP Paper: https://arxiv.org/pdf/2103.17249.pdf

StyleCLIP Code: https://github.com/orpatashnik/StyleCLIP